AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently Amended) A system that facilitates enhancement of a speech signal, comprising:

an input component that receives a speech signal and pixel-based image data relating to an originator of the speech signal; and[[,]]

a speech enhancement component that ~~employs a probabilistic-based model that~~ is configured to infer correlat[[es]]ions between the speech signal and the pixel-based image data ~~so as to facilitate discrimination of noise from the speech signal, the~~ by employing a probabilistic-based model comprising a video embedded subspace model fused with an audio mixture model such that ~~employing~~ a ~~set of~~ hidden variable[[s]] that represents the pixel-based image data in lower dimensions depends on a state variable of ~~representing relevant features, the features being inferred from at least one of~~ the speech signal ~~and pixel-based image data~~.

2. (Currently Amended) The system of claim 1, ~~the probabilistic-based model comprising an audio model,~~ wherein the audio model is based[[,]] at least in part[[,]] upon:

$$p(u \mid s) = \prod_k N(u_k \mid 0, \sigma_{sk})$$

$$p(s) = \pi_s$$

$$p(w \mid u) = \prod_k N(w_k \mid hu_k, \phi_k)$$

where u_k is a clean speech signal,

w_k is the speech signal,

s is [[a]]the state variable of the speech signal, and[[,]]
the notation $N(x \mid \mu, \sigma)$ denotes a Gaussian distribution over random variable x with mean μ and inverse covariance σ.

3. (Currently Amended) The system of claim 1, ~~the probabilistic-based model comprising a video model,~~ wherein the video model is based[[,]] at least in part[[,]] upon:

$$p(l) = const.$$
$$p(v \mid r) = \prod_i N(v_i \mid \sum_j A_{ij} r_j + \mu_i, \nu_i)$$
$$p(y \mid v, l) = \prod_i N(y_i \mid v_{i-l}, \lambda)$$

where y is the pixel-based image data,
r is [[a]]the hidden variable that represents the pixel-based image data in lower dimensions,
A is a matrix of weights for the hidden variable[[s]] r,
l is a location parameter,
v is a hidden clean pixel-based image,
v_{i-l} is shorthand for $v_\xi(x_i - x_l)$,
$x(i)$ is the position of the i^{th} pixel,
x_l is the position represented by l, and[[,]]
$\xi(x)$ is the index of v corresponding to 2D position x.

4. (Currently Amended) The system of claim 1, wherein the probabilistic-based model ~~comprising an audio/video model, the audio/video model~~ is based[[,]] at least in part[[,]] upon:

$$p(r \mid s) = \prod_j N(r_j \mid \eta_{sj}, \psi_{sj})$$

where r is [[a]]the hidden variable that represents the pixel-based image data in lower dimensions,

s is [[a]]the state variable of the speech signal,

ψ is a precision matrix parameter associated with s, and[[,]]

η is a precision matrix parameter associated with s.

5. (Currently Amended) The system of claim 1, wherein the speech enhancement component is configured to infer the correlations between the speech signal and the pixel-based image data ~~modification of at least one parameter of the probabilistic model being~~ based upon a variational expectation maximization algorithm having an E-step and an M-step.

6. (Currently Amended) The system of claim 5, wherein the variational expectation maximization algorithm ~~being~~ is based[[,]] at least in part[[,]] on ~~the equation~~:

$$p(u,s,r,v \mid y,w) \approx q(u \mid s)q(s)q(r \mid s)q(v \mid r,l)q(l)$$

where u is a clean speech signal,

s is [[a]]the state variable of the speech signal,

r is [[a]]the hidden variable that represents the pixel-based image data in lower dimensions,

v is a hidden clean pixel-based image,

y is the pixel-based image,

w is the speech signal, and[[,]]

l is a location parameter.

7. (Currently Amended) The system of claim 5, wherein the expectation maximization algorithm ~~being~~ is based[[,]] at least in part[[,]] on ~~the equation~~:

$$h = \frac{\mathrm{Re}\sum_k \phi_k \langle wkEu_k^* \rangle}{\sum_k \phi_k \langle E \mid u_k \mid^2 \rangle}$$

$$\frac{1}{\phi_k} = \left\langle \mid w_k \mid^2 \right\rangle - 2h\,\mathrm{Re}\left\langle w_k E u_k^* \right\rangle + \left\langle E \mid u_k \mid^2 \right\rangle$$

where

$$Eu_k = \sum_s \overline{\pi}_s \overline{\rho}_{sk}$$

$$E \mid u_k \mid^2 = \sum_s \overline{\pi}_s \left(\mid \overline{\rho}_{sk} \mid^2 + \frac{1}{\overline{\sigma}_{sk}} \right)$$

and[[,]]

u_k is a clean speech signal,

w_k is the speech signal,

π_s is a prior probability parameter of s, and

σ_{sk} is an inverse covariance~~, and,~~.

8. (Currently Amended) The system of claim 7, wherein the expectation maximization algorithm ~~being~~ is further based[[,]] at least in part[[,]] on ~~the equation~~:

$$A = \left\langle Evr^T - EvEr^T \right\rangle \left\langle Err^T - ErEr^T \right\rangle^{-1}$$

$$\mu = \left\langle Ev - AEr \right\rangle$$

$$v^{-1} = Diag\left\langle Evv^T - AErv^T - \mu Ev^T \right\rangle$$

where "Diag" refers to [[the]]a diagonal of the matrix, and[[,]]

$$Er = \sum_s \overline{\pi}_s \overline{\eta}_s$$

$$Err^T = \sum_s \overline{\pi}_s \left(\overline{\eta}_s \overline{\eta}_s^T + \overline{\psi}_s^{-1} \right)$$

$$
\begin{aligned}
Ev &= \sum_s \overline{\pi}_s \left(\overline{A}\,\overline{\eta}_s + \overline{\mu} \right) \\
Evr^T &= \sum_s \overline{\pi}_s \left[\left(\overline{A}\,\overline{\eta}_s + \overline{\mu} \right) \overline{\eta}_s^T + \overline{A}\,\overline{\psi}_s^{-1} \right] \\
Evv^T &= \sum_s \overline{\pi}_s \left[\left(\overline{A}\,\overline{\eta}_s + \overline{\mu} \right) \left(\overline{A}\,\overline{\eta}_s + \overline{\mu} \right)^T + \overline{A}\,\overline{\psi}_s^{-1} \overline{A}^T + \overline{\nu}^{-1} \right]
\end{aligned}
$$

9. (Currently Amended) The system of claim 8, wherein the expectation maximization algorithm ~~being~~ is further based[[,]] at least in part[[,]] on the equation:

$$
\begin{aligned}
\eta_{sj} &= \left\langle \overline{\eta}_{sj} \right\rangle \\
\frac{1}{\psi_{sj}} &= \left\langle \left(\overline{\eta}_{sj} - \eta_{sj} \right)^2 + \left(\psi_s^{-1} \right)_{jj} \right\rangle
\end{aligned}
$$

10. (Currently Amended) The system of claim 1, wherein the pixel-based image data compris[[ing]]es information associated with an appearance of ~~the~~ lips of the originator of the speech signal.

11. (Currently Amended) The system of claim 1, wherein the speech enhancement component that is configured to infer correlations between the speech signal and the pixel-based image data comprises a speech component that is configured to track[[s]] [[the]] lips of the originator of the speech signal in order to facilitate discrimination of noise from the speech signal.

12. (Currently Amended) The system of claim 1, wherein the input component further compris[[ing]]es a frequency transformation component that is configured to receive[[s]] windowed signal inputs, compute[[s]] a frequency transform of the windowed signal[[s]] inputs, and provide[[s]] outputs of the frequency transformed windowed signal[[s]] inputs to the speech enhancement component.

13. (Currently Amended) The system of claim 12, further comprising a windowing component that is configured to appl[[ies]]y an N-point window to the speech signal and provide[[s]] [[the]] windowed signal inputs to the frequency transformation component.

14. (Currently Amended) The system of claim 1, further comprising at least two audio input devices that is configured to provide speech signals.

15. (Currently Amended) The system of claim 1, wherein the probabilistic-based model is configured to be[[ing]] trained[[,]] at least in part[[,]] during operation of the system.

16-17. (Canceled).

18. (Currently Amended) A method of facilitating enhancement of a speech signal, comprising:

receiving a speech signal;

receiving [[a]] pixel-based image data relating to an originator of the speech signal; ~~and,~~

inferring correlations between the speech signal and the pixel-based image data using a probabilistic-based model comprising a video embedded subspace model fused with an audio mixture model such that a hidden variable that represents the pixel-based image data in lower dimensions depends on a state variable of the speech signal; and

generating an enhanced speech signal based[[,]] at least in part[[,]] upon ~~a probabilistic-based model that~~ the correlat[[es]]ions between the speech signal and the pixel-based image data ~~so as to facilitate discrimination of noise from the speech signal~~.

19. (Original) The method of claim 18 further comprising providing an output associated with the enhanced speech signal.

20. (Currently Amended) A data packet configured to be transmitted between two or more computer components that are configured to facilitate[[s]] enhancement of a speech signal, the data packet comprising:

an enhanced speech signal~~, the enhanced speech signal being based,~~ generated at least in part[[,]] ~~upon~~ utilizing a probabilistic-based model that is configured to infer correlat[[es]]ions between a speech signal and image data related to an originator of the speech signal, the probabilistic-based model comprising a video embedded subspace model fused with an audio mixture model such that a hidden variable that represents the image data in lower dimensions depends on a state variable of the speech signal~~so as to facilitate discrimination of noise from the speech signal~~.

21. (Currently Amended) A computer readable medium storing computer executable components of a system that facilitates enhancement of a speech signal ~~comprising~~, the computer executable components comprising:

an input component ~~that~~ configured to receive[[s]] a speech signal and ~~pixel-based~~ image data relating to an originator of the speech signal; and[[,]]

an speech enhancement component ~~that~~ configured to employ[[s]] a probabilistic-based model that is configured to correlate[[s]] ~~between~~ the speech signal and the image data, the probabilistic-based model comprising a video embedded subspace model fused with an audio mixture model such that a hidden variable that represents the image data in lower dimensions depends on a state variable of the speech signal~~so as to facilitate discrimination of noise from the speech signal~~.

22. (Currently Amended) A system that facilitates enhancement of a speech signal comprising:

means for receiving a speech signal and ~~pixel-based~~ image data relating to an originator of the speech signal; and[[,]]

means for enhancing the speech signal, the means for enhancing configured to employ[[ing]] a probabilistic-based model that is configured to correlate[[s]] ~~between~~ the speech signal and the image data, the probabilistic-based model comprising a video embedded subspace model fused with an audio mixture model such that a hidden variable that represents the image data in lower dimensions depends on a state variable of the speech signal~~so as to facilitate discrimination of noise from the speech signal~~.